

Data management workflows for campaigns and model data

Linda Baldewein

Helmholtz Coastal Data Center
Institute of Carbon Cycles
Helmholtz-Zentrum Hereon
Geesthacht, 29.09.2021

DAM



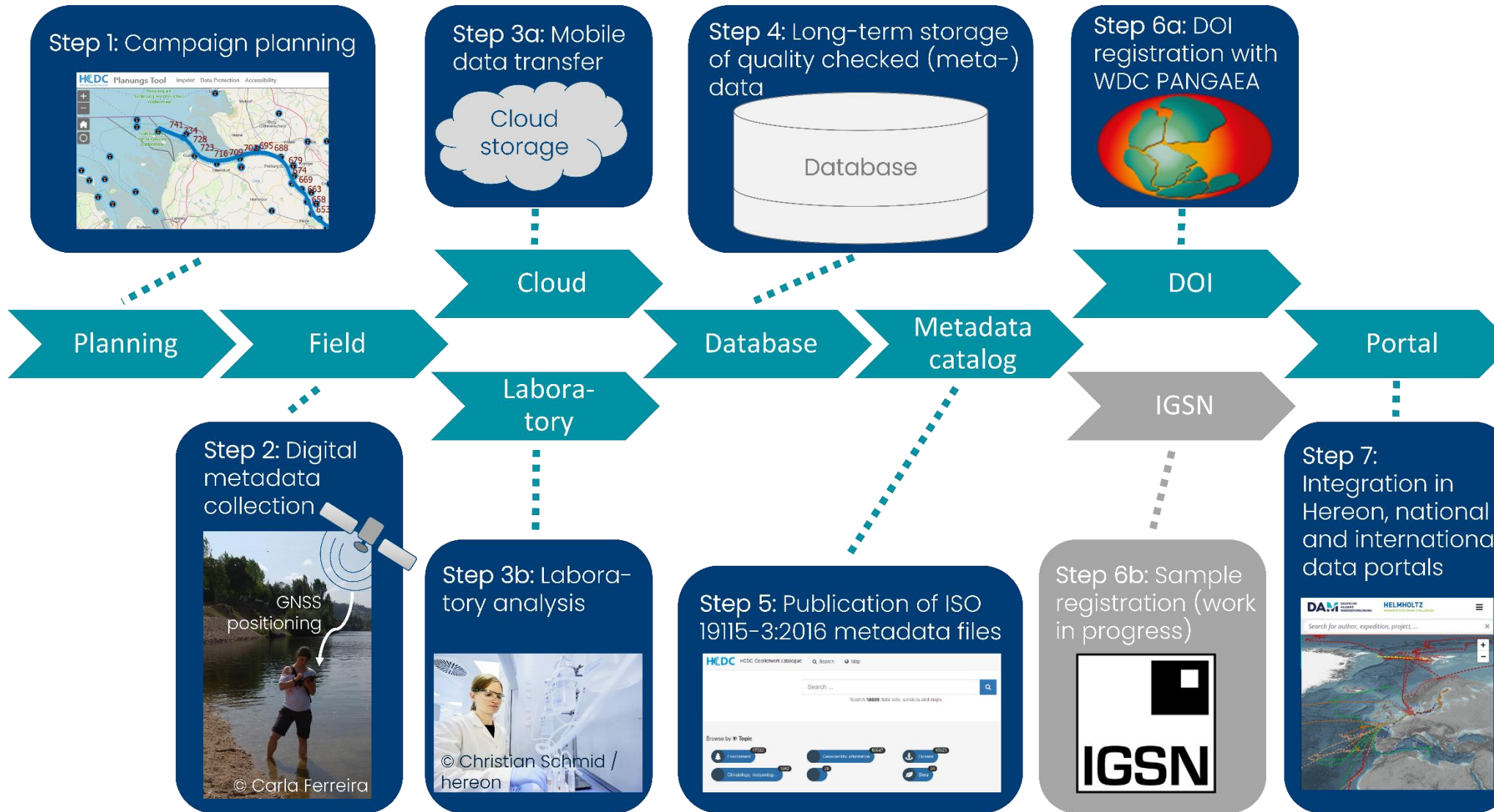
Helmholtz-Zentrum
hereon

Agenda

- Example workflow for campaigns
 - Hereon campaign data workflow
- Example workflow for models
 - Introduction to the Model Data Explorer



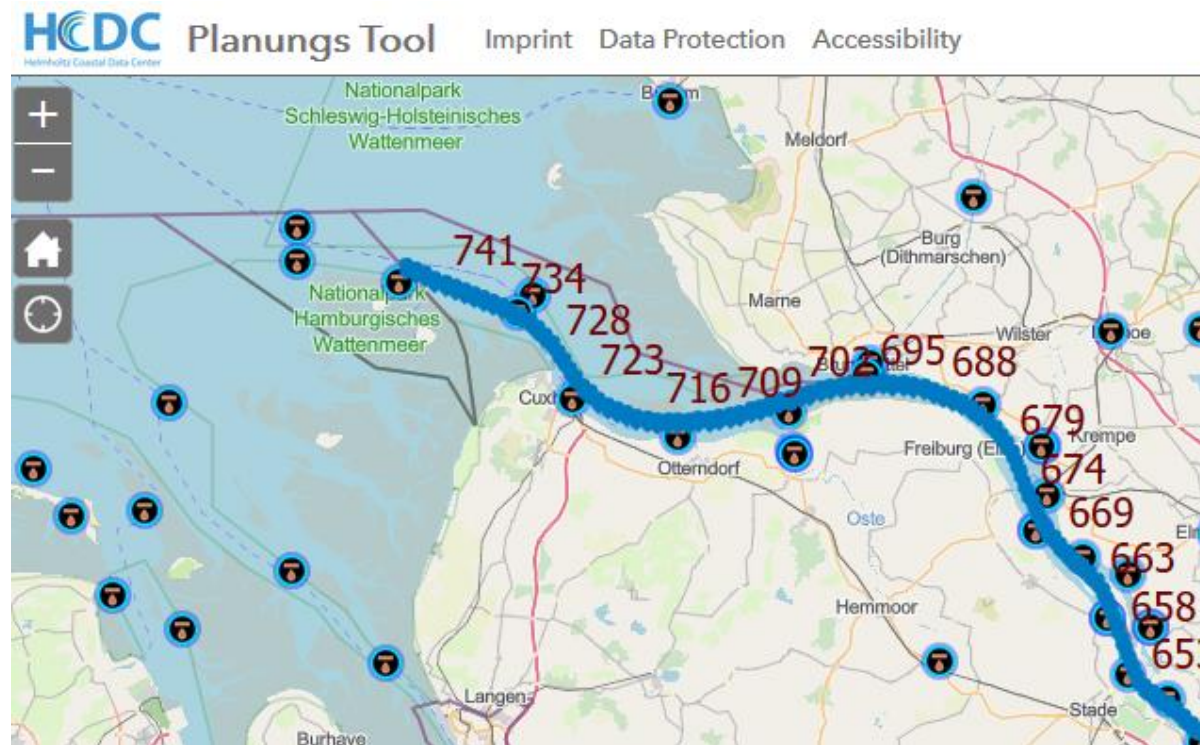
Campaign data workflow



Campaign data: Planning

Planning

- Where will the campaign take place?
- How many events / stations are planned?
- Create planned event / station list and / or web map
 - Most German research institutions have an ArcGIS Online subscription without you knowing.
 - QGIS as open-source Desktop alternative



Campaign data: Planning

Planning

- How will metadata be recorded?
 - DSHIP on larger research vessels
 - Handwritten
 - Survey123 App
- Who will record metadata?

Science Activity Number	Science Activity Comment	Science Activity Closed	Science Activity Area
Comment	Device Operation	Device Operation	Device Operation
Closed Device Operation Label	Device Shortname	Device Comment	Underway
Comment	Action	Latitude	Longitude
(deg)	Wind Dir	Wind Velocity	Event Time
0	f	1	AL557_0_Underway-1
ALKOR	04/06/2021 06:54	profile start	"53° 57,841' N" "008° 37,109' E"
312.6	53.964012	8.618485	101 2.8
1	t	1	AL557_1-1
the water	"53° 57,749' N"	"008° 37,225' E"	15 1 117.1
1	t	1	AL557_1-1
deck "53° 57,661' N"	"008° 37,406' E"	16 0.6 143.2	53.96101
2	t	1	AL557_2-1
the water	"54° 00,007' N"	"008° 05,928' E"	24 0.2 62.6
2	t	1	AL557_2-1
deck "54° 00,020' N"	"008° 05,939' E"	24 0.1 324.2	54.000335
3	t	1	AL557_3-1
the water	"54° 03,665' N"	"008° 01,021' E"	27 0.4 265.5
3	t	1	AL557_3-1
deck "54° 03,674' N"	"008° 00,917' E"	27 0.1 283.8	54.061235
3	t	2	AL557_3-2
04/06/2021 10:43	in the water	"54° 03,673' N"	"008° 00,922' E"
54.061222	8.015373	290.2	0.8

Campaign data: Collecting

Required information

- Metadata of your campaign
 - Ship name
 - Station / Event name
 - Coordinates
 - Sample information
 - Participants

-> Ideally digitally collected using DSHIP or Survey123



Campaign data: Processing and Analysing

- Metadata of your measurement
 - Parameter name
 - Unit
 - Method
 - Quality information
 - Origin of sample

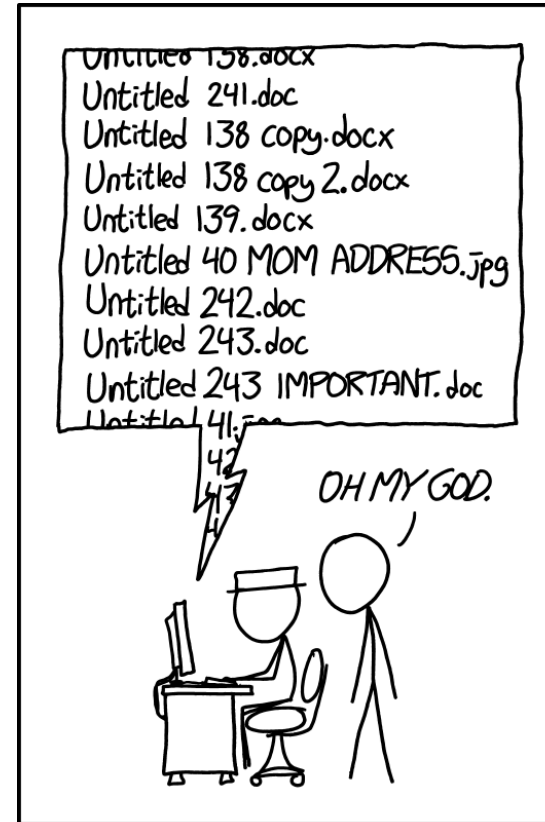
Processing
and
Analysing



© Christian Schmid / Hereon

Campaign data: Preserving

- Store results locally
 - Meaningful file names
 - Shared storage with your working group
 - Include metadata
- Prepare for submission
 - Format based on requirements of repository
 - Use template, if available
 - Double check for typos, etc.
- Submit to institution repository (if available)



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Randall Munroe, 2014, CC BY-NC 2.5

Campaign data: Publishing and Sharing

Publishing
and
Sharing

- Publish your data
 - DOI registration with PANGAEA
 - See details in talk by Flavia Höring
- Publish your samples
 - IGSN
- Your local data curators may help you. Ask them!

GFZ
Helmholtz Centre
POTSDAM

IGSN
Physical Sample

General Identifiers

Project:	Project 1 and 11
Campaign:	Litter harvest Exp 2017
Type:	Specimen
Name:	Senna cumingi
IGSN:	GFRCH001G
Parent IGSN:	N/A
Release Date:	N/A

Sampling Location

Latitude:	-30.053975
Longitude:	-71.0950333
Coordinate System:	WGS84
Elevation:	629.0
Location Type:	N/A
Location Name:	QT Plot 1
Location Description:	10x10 m Plot, established by Rafaela Canessa in April 2016. The study site is called Quebrada de Talca, and corresponds to a private conservation project, with past cattle grazing use and subsequent animal exclusion (for 7 years in 2016). Quebrada de Talca presents similar vegetation, parent material and climatic conditions to Santa Gracia, but thanks to the exclusion project, the vegetation grows better and has no current animal influence.
Country:	Chile
Province:	Coquimbo Region
County:	N/A
City:	N/A

Acquisition

Material:	Biology
Biology Classification:	vegetation:leaf litter
Biology Description:	5 glasses with milled litter Nr: 725, 726, 727, 728, 729.
Collection Method:	manual
Field Number:	

Sample Family

Sample Family shows a sub-sampling graph. Select entries to navigate samples.

Location Map

Drilling Start/End: 2017-01-17 / 2017-01-17 *
Latitude: -30.05398 * Longitude: -71.09503 *
QT Plot 1

PANGAEA.
Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Citation:

Apel, Christina; Joerss, Hanna; Ebinghaus, Ralf (2018): Organic UV stabilizers and UV filters in the sediment of European North and Baltic Seas in 2016/2017. *PANGAEA*, <https://doi.org/10.1594/PANGAEA.895397>,
Supplement to: Apel, C et al. (2018): Environmental occurrence and hazard of organic UV stabilizers and UV filters in the sediment of European North and Baltic Seas. Chemosphere, 212, 254-261, https://doi.org/10.1016/j.chemosphere.2018.08.105

Always quote above citation when using data! You can download the citation in several formats below.

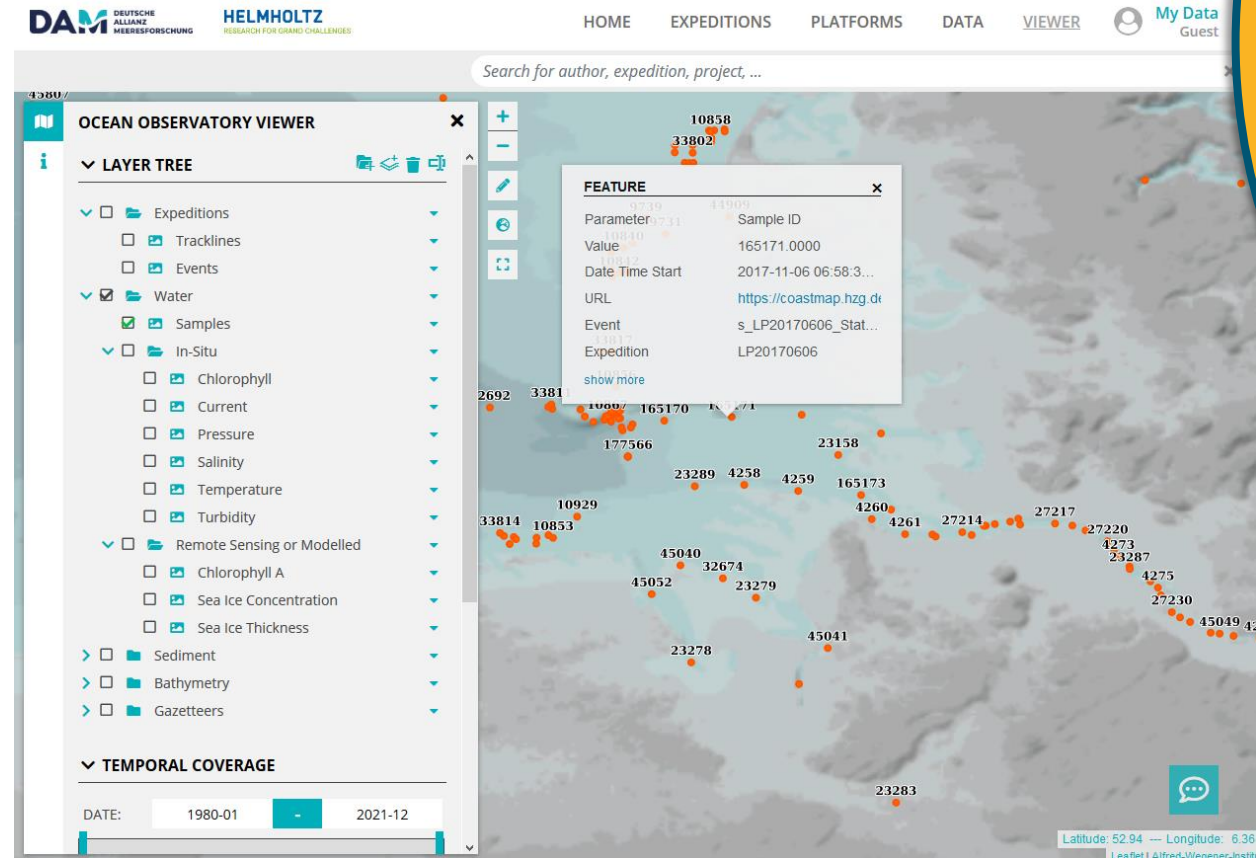
RES Citation BibTeX Citation Copy Citation Facebook Twitter Show Map Google Earth

Abstract:

UV absorbing compounds are of emerging concern due to their large production volumes, their persistence or pseudo-persistence, and their potential adverse effects. This is the first study investigating the environmental occurrence and potential hazard of organic UV stabilizers and UV filters in the North and Baltic Sea surface sediments, including the connecting Skagerrak and Kattegat straits. In total, nineteen substances were identified over the entire study area, including the rarely studied compounds ethylhexyl triazone (EHT) and bisocryzole (UV-360). Octocrylene (OC) was the predominant compound in this study with regard to detection frequency (79%) and concentrations (up to 9.7 ng/g dwt). OC accounted for more than 80% of UV stabilizer concentration in the German Bight. The relative detection of EHT was identified in the

Campaign data: Reusing

- Data available in data portals
- Other scientists can find, download and cite your data



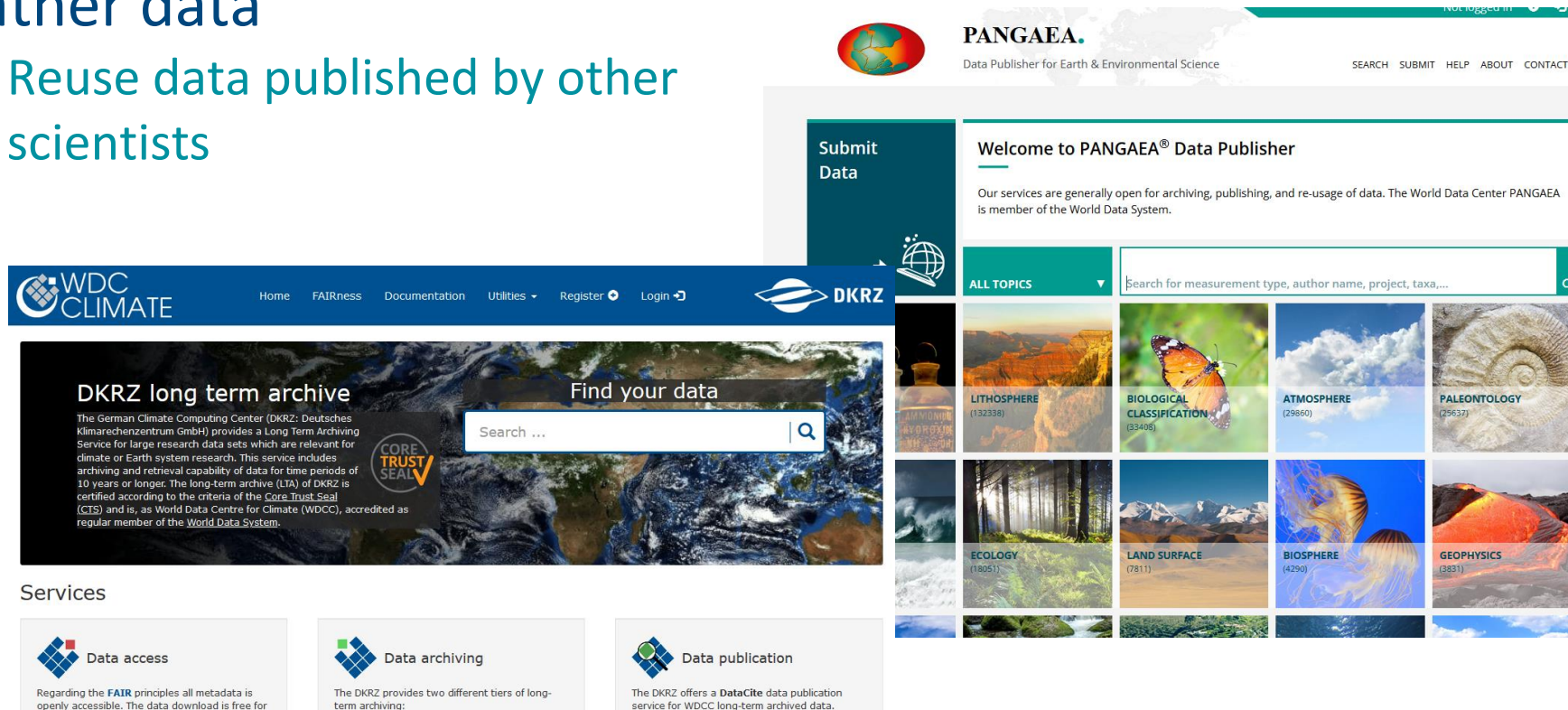
Reusing

Model data: Planning and Collecting

- Which model will you use / develop?
- Gather data
 - Reuse data published by other scientists

Planning

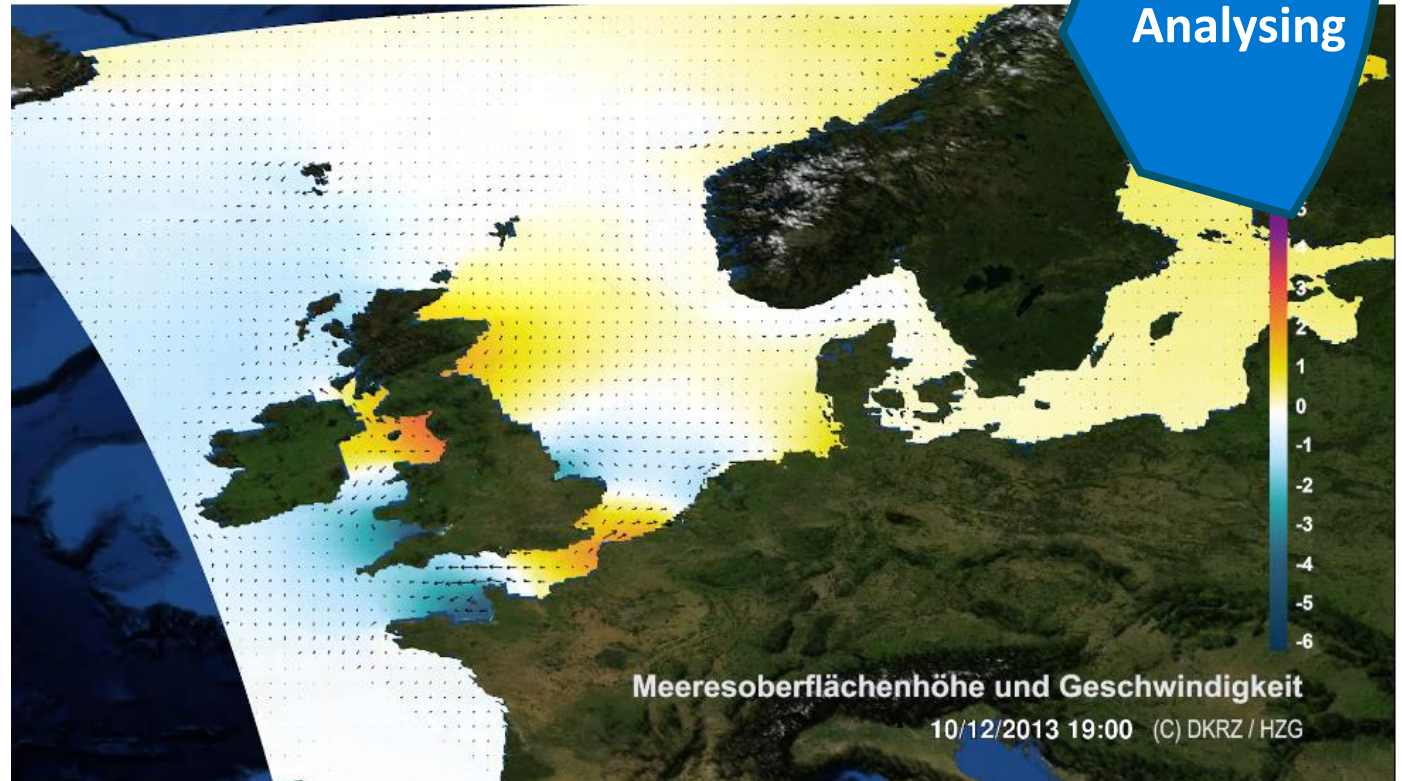
Collecting



The image displays two screenshots of scientific data repositories. The top screenshot is the PANGAEA website, which serves as a data publisher for Earth and Environmental Science. It features a search bar, a 'Submit Data' button, and a grid of categories including Lithosphere, Biological Classification, Atmosphere, Paleontology, Ecology, Land Surface, Biosphere, and Geophysics. The bottom screenshot is the DKRZ website, part of the WDC Climate network. It highlights the 'DKRZ long term archive' and provides a search interface. Below the main content, there are three service boxes: 'Data access' (regarding FAIR principles), 'Data archiving' (offering two tiers of long-term archiving), and 'Data publication' (offering a DataCite service for WDC long-term archived data).

Model data: Processing and Analysing


- Metadata of your model
 - Conventions
 - Institution, originator, contact
 - Title
 - Source
 - Creation date
 - Coordinate reference system
 - Dimensions
 - Variables
 - Units
 - Standard names



Model data: Preserving

- Store results locally
 - Meaningful file names
 - Shared storage with your working group
 - Include metadata
- Prepare for submission
 - Format based on requirements of repository
 - Use template, if available
 - Double check for typos, etc.
- General rules for data storage in netCDF files compiled at Hereon



	Hereon Data Management	Name: Binding Regulations for Storing Data as netCDF Files Date: 18.05.2021
---	---------------------------	---

2. General Specifications

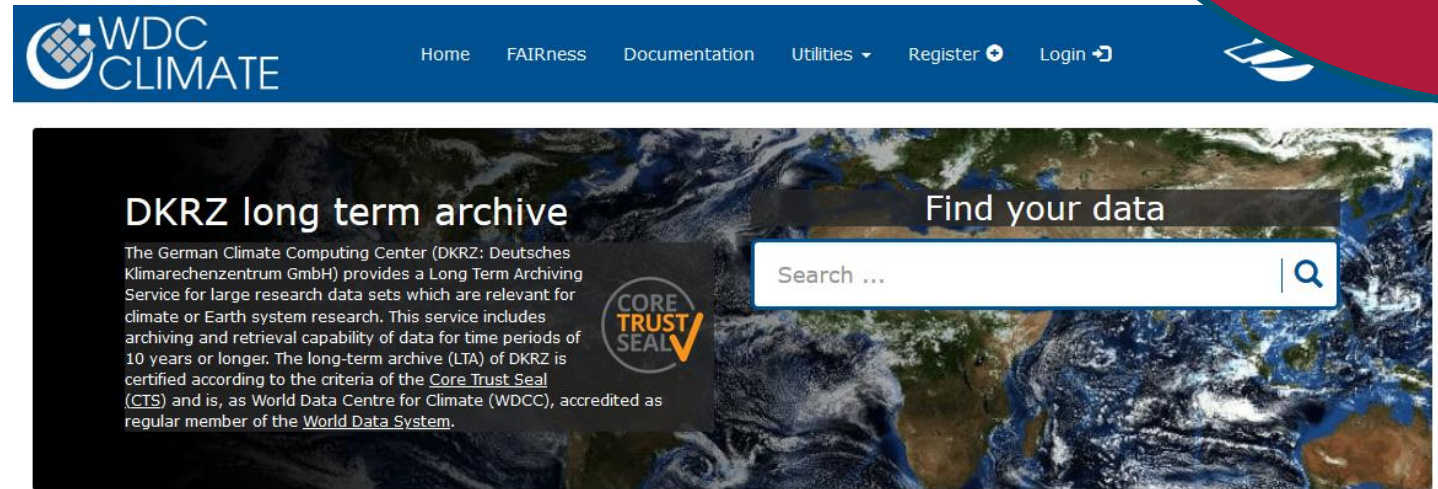
Provided here is general information that has nothing directly to do with the netCDF format itself but should, however, be taken into consideration.

2.1. File Names

- NetCDF files are designated with the extension .nc
 - The file name begins with a letter or a letter sequence. The letter sequence should allow users to draw conclusions about the type of data found in the file.
 - Customary sequences used thus far (e.g., "ctd" for CTD data, "sf" for ScanFish data) will be retained. Specification should still be designated for other devices.
 - When dealing with model data, either the model name with the version designation or the ExperimentID should be used in the first segment of the filename.
- Spaces, umlauts, "ß", special characters (except underscore "_", dash "-", and periods) are not

Model data: Publishing and Sharing

- Publish your data
 - DOI registration with WDC CERA at DKRZ
- Your local data curators may help you. Ask them!

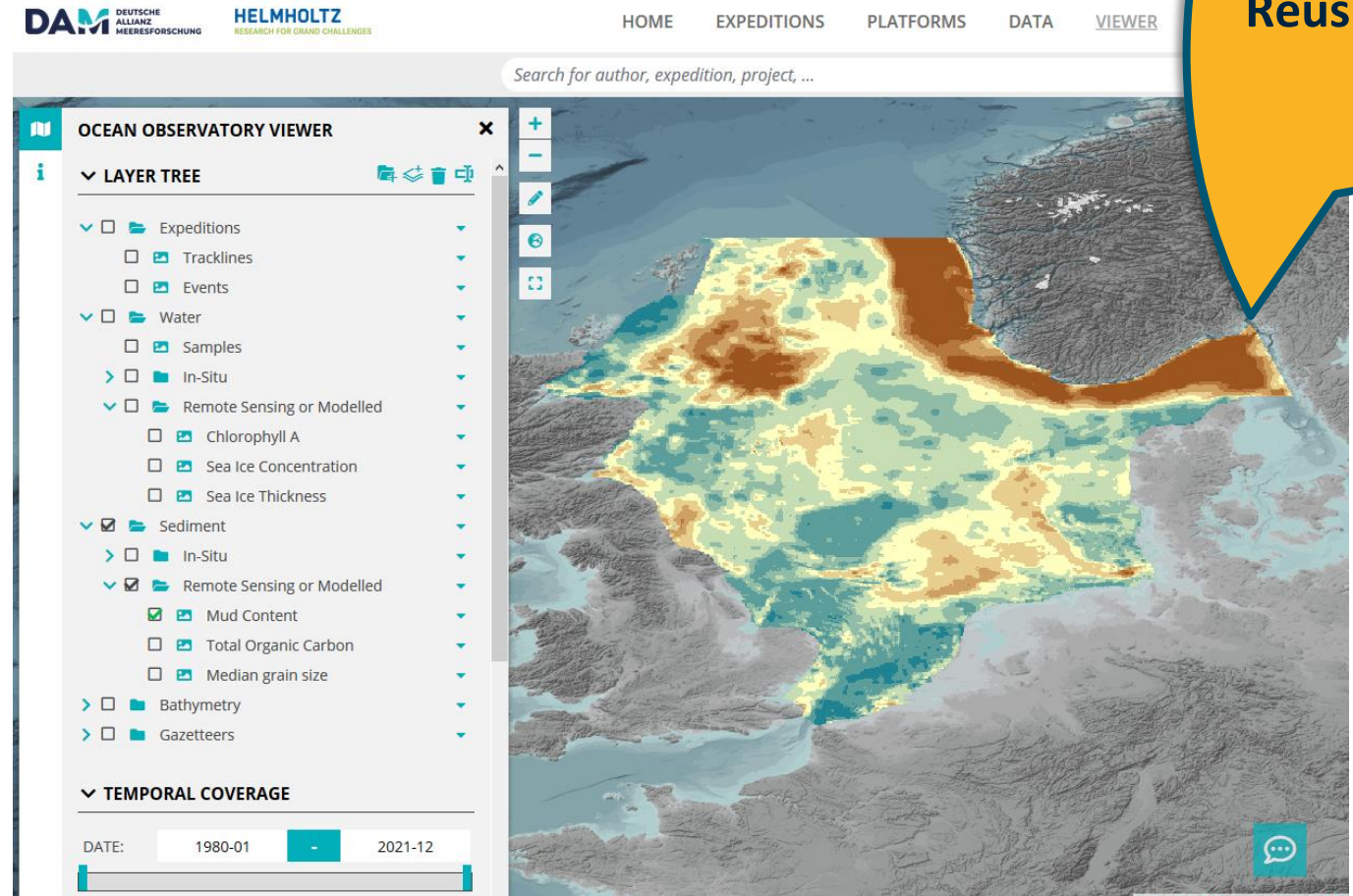


Services



Model data: Reusing

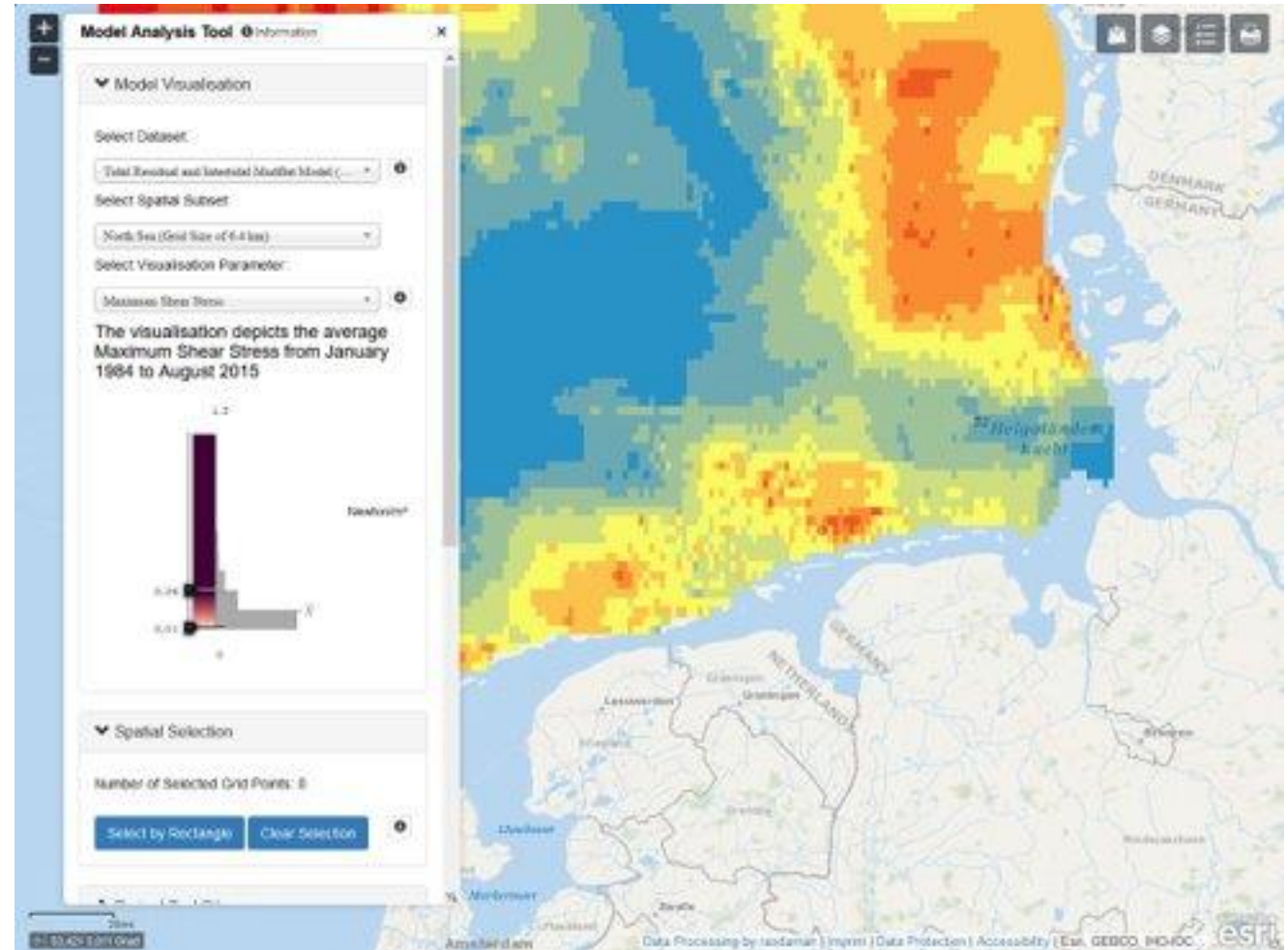
- Other scientists can find, download and cite your data
- Data available in data portals
 - Currently a lot of manual effort by curators



Reusing

Model Data Explorer

- Central platform to access Model Data
 - 4D Model Data Map
 - Search and filter datasets
 - Compute and compare statistics on the data
 - Download raw data
- Currently in development
 - We need support from future users
 - Join our mailing list mde-dev@listserv.dfn.de
 - Check out the [development](#)



Vielen Dank.

www.hereon.de

